2014

Traffic Pattern Detection Using the Hough Transformation for Anomaly Detection to Improve Maritime Domain Awareness

McAbee, Ashley; Scrofani, James; Tummala, Murali; Garren, David; McEachen, John

IEEE

McAbee, Ashley, James Scrofani, Murali Tummala, David Garren, and John McEachen. "Traffic pattern detection using the Hough transformation for anomaly detection to improve maritime domain awareness." In Information Fusion (FUSION), 2014 17th International Conference on, pp. 1-6. IEEE, 2014.

http://hdl.handle.net/10945/59167

Traffic Pattern Detection Using the Hough Transformation for Anomaly Detection to Improve Maritime Domain Awareness

Ashley McAbee, James Scrofani, Murali Tummala, David Garren, and John McEachen
Electrical and Computer Engineering Department
Naval Postgraduate School
Monterey, California, USA
jwscrofa@nps.edu, mtummala@nps.edu

***Abstract*—Techniques for extracting traffic patterns from ship position data to generate atlases of expected ocean travel are developed in this paper. An archive of historical data is used to develop a traffic density grid. The Hough transformation is used to extract linear patterns of elevated density from the traffic density grid, which can be considered the "highways" of the oceans. These highways collectively create an atlas that is used to define geographical regions of expected ship locations. The atlas generation techniques are demonstrated using automated information system (AIS) ship position data to detect highways in both open-ocean and coastal areas. Additionally, the atlas generation techniques are used to explore variability in ship traffic as a result of extreme weather. The development of an automatic atlas generation technique that can be used to develop a definition of normal maritime behavior is a significant result of this research.**

***Keywords*—Maritime Domain Awareness, Hough Transformation, Automated Information System, Pattern Extraction**

I. Introduction

Data alone are not enough to achieve a useful common operating picture in the maritime domain; an ability to identify trends and differentiate anomalies is required for maritime domain awareness (MDA) objectives to be met. There were more than 55,000 port calls in 2009 from nearly 7,000 different oceangoing vessels in the United States alone [1]. Rapid decision making and response in this busy environment requires automated methods to turn the volumes of raw data collected on these vessels into processed intelligence. The research detailed in this paper is a development of techniques that automate pattern extraction from large volumes of maritime vessel position data to establish normal behavior and thereby enable anomaly detection.

The objective of the research detailed in this paper is to contribute to MDA by enhancing anomaly detection through the identification of normal traffic patterns extracted from archived ship position data. The extracted patterns, or highways, are determined using the Hough transformation technique and are compiled into atlases of expected ocean traffic patterns that capture a geographical-position based definition of normal maritime behavior.

The human eye is capable of seeing how features are arranged in images, while computational algorithms are necessary to automate the same discernment [2]. The techniques proposed in this paper use the Hough transformation methods outlined in [2] and [3] and tailor those methods to accommodate data related to maritime vessel traffic so that the computer can extract the linear traffic patterns, or highways. Those extracted highways taken collectively into an atlas provide a template of normal vessel behavior to be used to comparatively identify anomalous behavior. The developed method was applied experimentally to an archive of historical data to demonstrate the concept of a complete anomaly detection architecture.

The pattern extraction and anomaly detection methods used in this paper rely heavily on having accurate position reporting concerning vessels of interest. While these position reports could come from any means, including radar, satellite imagery, or observations recorded from trusted vessels, the development of automated information system (AIS) has created large archives of ship position reports that lend themselves well to research and development of MDA tools. The pattern extraction techniques could be used with any data source and more powerfully still with a fusion of multisource intelligence data sources, but only AIS data are used in this paper as a proof of concept.

The AIS was developed to provide ship operators with integrated displays of all ships within their very high frequency radio range. It was conceived as a mechanism for improving safety at sea by enabling ships to clearly identify other ships around them, not just by location on a radar screen, but also by specific name [4]. In the U.S. navigable waters, the system is now required by US Coast Guard regulation on all passenger vessels of more than 150 gross tons displacement or certified to carry more than 150 passengers-for-hire, all tankers, all vessels of more than 300 gross tons displacement, all sail boats over 65 feet in length, and all towing vessels over 26 feet in length [4].

This paper presents a method for extracting traffic patterns from ship position data to generate atlases of expected ocean travel utilizing historical AIS data. The proposed method is used to explore variability in ship traffic as a result of extreme weather.

The paper is organized as follows. Section II presents the details of the proposed method for determining the traffic patterns based on the Hough transformation to obtain the atlases. The results demonstrating the ability of the proposed method for atlas generation utilizing the AIS data are included in Section III. Changes in traffic patterns as a result of Hurricane Ernesto in the Caribbean are illustrated.

II. PROPOSED METHOD

Archives of data on ship tracks provide a wealth of information from which expected ship behavior over a variety of scenarios can be discerned. A common perception in literature related to MDA is that vessel tracks are preserved throughout the anomaly detection process. As an alternative technique, an image processing technique is used in this paper to develop an atlas of ship motion based on statistical analysis of position reports with no need for unique vessel identification. The nine distinct steps depicted in Figure 1 outline the approach taken in this paper to develop an atlas of expected ocean highways from archives of ship position reports.

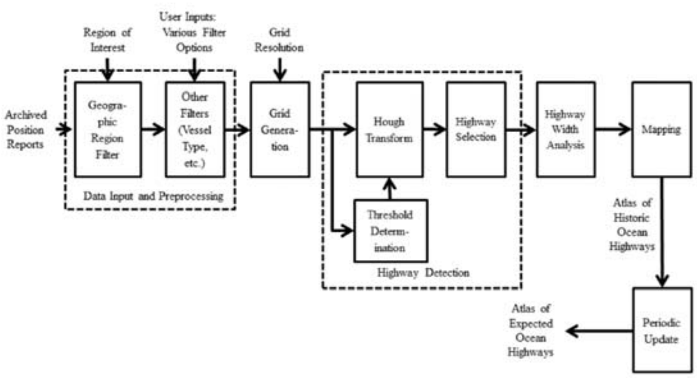

Figure 1. The atlas of historic ocean highways is tailored to a user's needs by taking input from the user on the region of interest, various filter options, and the required resolution of the area.

A. Data Input and Preprocessing

Preprocessing involves accepting data in the sensor output format and filtering and preparing that data so that follow-on steps can be applied. This is the first step in the flow process outlined in Figure 1. The user selects the region of interest (ROI). From the complete archive of position reports G, the data is filtered to include only the ROI as given by

$$G_{ROI} = G\{long_{\min} < G_x < long_{\max};\ lat_{\min} < G_y < lat_{\max}\}$$

where G_{ROI} denotes the data only from the ROI, G_x and G_y represent the coordinates of each position report in longitude and latitude, and *long* and *lat* min and max represent the user defined boundaries to the ROI. The available archive of data is filtered first by timestamp to the window of interest and then by latitude and longitude to the geographic area of interest. Data are then further filtered as necessary for other details, such as vessel type or flag of origin, depending on user preferences.

B. Grid Generation

Grid generation is next as we continue through the flow diagram outlined in Figure 1. The user determines the grid resolution that should be employed. The highway determination method in follow-on steps uses the relative density of position reports in disparate sections of the ROI to detect the most frequently traveled routes in that area. Thus, the ROI must be divided into a grid with an equal number of rows and columns that can be used to develop a count of how many position reports occur in each region of the grid. The mesh size for this grid is determined by the user because the same grid size is not applicable in every scenario. The mesh size refers to the size of each region of the grid along the axis of longitude in minutes.

Many factors impact what the "right" value is for a given region [5]. As in most digital applications, increasing resolution is generally desirable; however, there are some limits in this application. First, the significant figures measured by the position report source set a lower limit of what resolution is possible. For example, position reports in AIS are only received to the fourth decimal place in minutes of longitude. Setting a mesh size less than this resolution results in mesh squares tied to position reports that can never exist in the data archive and creates false zero counts in the water space. Second, finer grids result in higher computational costs as the number of individual regions of the grid that must be considered grows. Third, less apparent associations may be overlooked. These missed highways are somewhat equivalent to the false negative problem in radar target analysis – a highway may be overlooked when it actually does exist. These points are illustrated in Figure 2. For example, if two ships are travelling on parallel courses with just 1.5 nautical miles (nm) of separation, or approximately 0.0250 minutes of longitude at the equator, and the grid is too fine, this may not be detected as a highway because of the zero traffic density regions between the two vessels. Selecting a mesh size that is too large can have negative impacts too. To begin, the grid size impacts the exactness of highway placement. Linear traffic patterns are detected from the grid, and the highway is overlaid across the associated grid squares. As a result, the placement has an inherent measurement accuracy of no more than +/- the diagonal length of a grid square. Further, nonexistent highways may be incidentally detected in a mesh that is too large. Equivalent to a false positive in the radar target analysis realm, this relates to a highway being detected that does not really exist.

C. *Highway Determination*

Once the grid has been developed, the Hough transformation [2] is used to detect highways, which is the third major section of the flow diagram in Figure 1.

1) *Traffic Density Threshold Determination*

Before the automatic detection of these co-linear densities can occur, a threshold for traffic densities of interest ρ_{TH} must be selected. The grid contains normalized traffic densities, so ρ_{TH} is some percentage of the maximum traffic density present in the ocean space. While a range of values can be selected, in this work the 75th percentile served as a reasonable starting point in most cases.

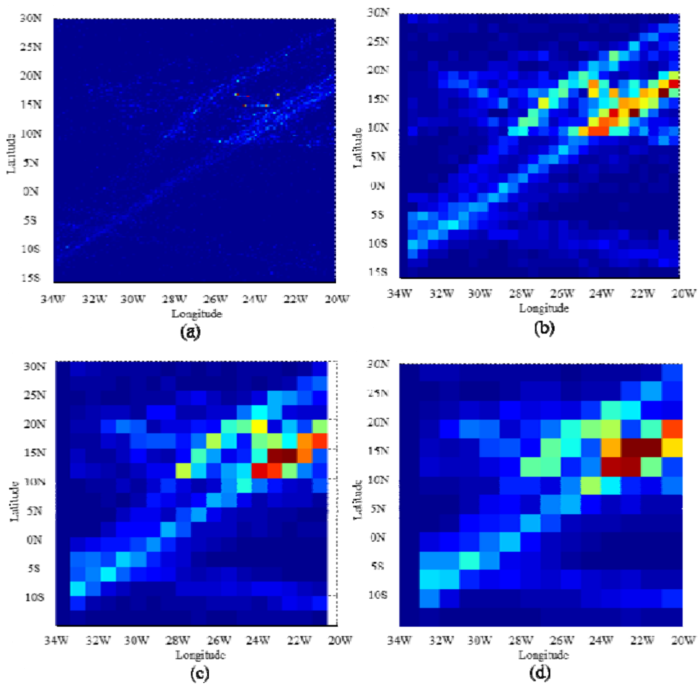

Figure 2. Comparing grid sizes (a) 0.10 minutes, (b) 0.50 minutes, (c) 0.75 minutes, and (d) 1.00 minutes provides insight into how to select the best size. Reductions in grid size increase processing time and result in lost highways, yet larger grid sizes reduce the accuracy of highway placement.

2) *Identification of the Dominant Trend*

Each region of the grid that meets or exceeds the traffic density threshold is transformed in the Hough space using the Hough transformation techniques described in [2],[5]. In the Hough space, each such grid region is defined by a set of values that includes the distance-angle pair (d, θ) of all possible straight lines through that grid region at indices (x, y). From the Hough space, co-linear regions are detected by identifying common (d, θ) pairs within each set S between different grid regions.

The Hough transformation is performed by translating a point in the reference axis system, or real space, to a set of definitions in the Hough space that exhaustively list the various straight lines that can be drawn through the point and how each of those lines relate to the origin of the reference axis system. Consider the points P_1 through P_5 presented in 0Every point P_n is transformed into a series of coordinates (d, θ) as illustrated in 0Every possible line $L_{\psi n}$ through P_n at an angle ψ with respect to a horizontal reference line is considered; five examples of these lines are depicted. For each $L_{\psi n}$, the closest point of approach D between the origin of the reference axis system and $L_{\psi n}$ is identified in polar coordinates (d, θ) where d is the distance from the origin to D, and θ is the angle between the horizontal axis through the origin and a line drawn from the origin to D.

The values of D_b, d_b, and θ_b are illustrated for P_3 and $L_{\psi b}$ in 0These values can be calculated by first defining Q to be a point that is one unit of distance away from P_n and along the same line that intersects P_n at angle ψ. The coordinates of Q will be

$$(Q_x, Q_y) = (P_{n,x} + \cos(\psi), P_{n,y} + \sin(\psi)).$$

Next, d can be found according to

$$d = \frac{\left|\det\begin{pmatrix} Q-P \\ O-P \end{pmatrix}\right|}{|Q-P|}$$

where O represents the origin located at (0, 0). The slope m can be found from

$$m = \frac{Q_y - P_y}{Q_x - P_x}$$

and the y-axis intercept b can be found from $P_y = mP_x + b$.

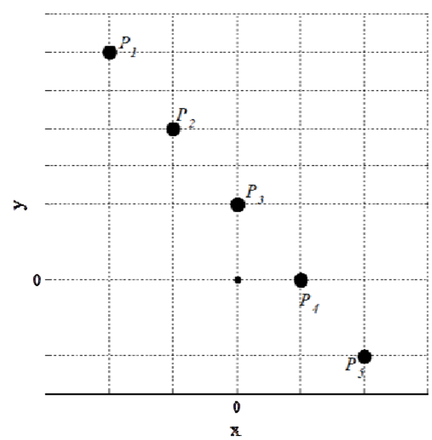

Figure 3. The Hough transformation to detect the co-linearity of points P_1 through P_5.

Since D must also lie along the same line as P and Q, the coordinates of that point can be found from the simultaneous solution to

$$D_y = mD_x + b$$

and

$$d = \sqrt{D_x^2 + D_y^2}$$

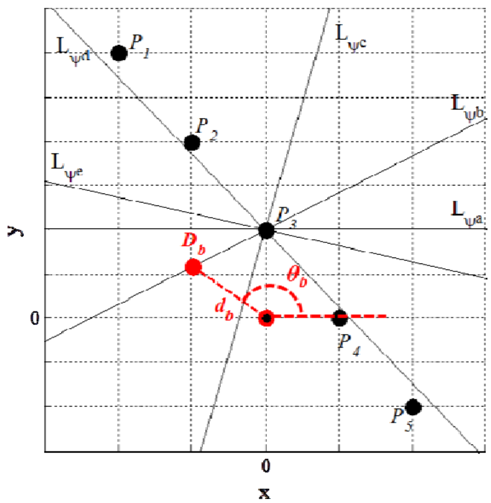

Figure 4. The distance to the closest point of approach D is found for every possible line through P_n.

which enables the determination of θ from

$$\theta = \cos^{-1}\left(\frac{D_x}{d}\right).$$

This process is repeated for $\psi = [1, 180^o]$ so that each point of interest P is expressed as a list of (d, θ) pairs, and this collection of (d, θ) pairs is often considered the Hough space.

As the term is used in this paper, the "dominant" trend refers to that linear region with the most co-linear grid regions that meet or exceed ρ_{TH}. The dominant trend pair (d_D, θ_D) can be found by taking into account the frequency of occurrence of (d, θ) across all grid regions. The most frequently occurring pair represents the dominant trend.

Although many applications directly use the straight-line definition derived from the Hough space to define the line detected in the real space, the maritime domain requires special consideration. Because highways may be of any width and, more specifically, because that width may change over the length of a highway, the direct result of the Hough space may not be the most comprehensive definition of the highway. Instead, a first-degree polynomial is fitted to the collection of all regions of the grid and a width study is performed to further define the highway.

3) Highway Width Analysis

Once the centerline of a highway is identified, a width study is performed to determine the upper and lower limits of the highway as seen from the process flow depicted in Figure 1. Terrestrial highways are relatively uniform in width, with changes occurring only as lanes are physically added to or removed from the infrastructure of the highway, but oceanic highways are not uniform since they are not generally physically bounded. As such, the width must be determined incrementally along the highway to best capture its variability. The algorithm determines the width at each incremental step along the highway by considering the densities of the regions of the grid along a line perpendicular to the detected highway at the given step. The highway width is located at the index of the last region of the grid along that line that still meets or exceeds a traffic density threshold ρ_w. The width is different on either side of the highway centerline and at each incremental step.

The threshold ρ_w can be set to any value, but one technique for determining it comes from finding the parameters of mean μ and standard deviation σ for the normal distribution curve fitting a profile of the traffic densities across the width of the highway at each incremental step. The density threshold ρ_w can then be set to define the width of the highway as a boundary within which all grid regions are within one standard deviation of the mean.

4) Trend Removal for Follow-On Detection

To enable the detection of less obvious linear trends in the data, the dominant trend is removed after it is discovered. Otherwise, secondary, tertiary, and further trends go undetected because they are masked. Simply removing the dominant trend and replacing those grid squares with zeros can unintentionally create new linear regions on the edges of the removed regions. Instead, the grid values that fall within a region bounded by the width definitions of the dominant trend are replaced by a pseudorandom value taken from the collection of all traffic density values occurring in the ROI. Once this replacement has occurred, the grid is renormalized to the new traffic density of interest, and the Hough transformation is performed again to identify the next-most-dominant trend in the region.

D. Mapping

One of the final steps outlined in the flow diagram in Figure 1 involves adding the highway to a map of the ROI. The Hough transformation is performed on a space indexed according to the reference grid. For the end results to be universally meaningful, they must be translated back into latitude and longitude. This is performed through linear interpolation based on how the grid was generated from the original latitude and longitude dimensions of the ROI.

III. RESULTS

The results presented in this paper were developed to exemplify the functionality of the highway determination method described in Section II, but selected case studies are not intended to draw conclusions regarding normal behavior of ships.

A. Demonstration of the complete method

The region of interest for this case study is the southern Atlantic Ocean. To ensure that only open-ocean traffic was observed, the region was bounded well outside of coastal waters. The exact region lies between the lines 15°S and 30°N in latitude and the lines of 30°W and 20°W in longitude. This area is outlined in black in Figure 5. The data set of AIS position reports was collected via satellite over seven months in 2012. A sample of the position reports received are plotted in Figure 5 in blue to provide a visual idea of how traffic is distributed in the area. The grid size is 0.2 minutes longitude. The Hough transformation is performed four times, with the recently identified trend plotted in red and removed after each iteration. Each iteration uses the new 75th percentile as the normalization factor, resulting in highway precedence as listed in Table 1 and identified in Figure 5. The general specifications of each of the highways are described in Table 1.

The propensity for maritime traffic to follow common routes, even on the unbounded open-ocean, can be discerned by considering the most dominant highway. The highway makes up just 17% of the total area of the region outlined in black in Figure 5, yet nearly half of all position reports received are located within the definition of this highway. In fact, 99.99% of all position reports received for the area fall on one of the four most dominant highways, even though they account for less than 50% of the total area of the region given the overlap between the first three highways. In seven months, only about 2,000 position reports were received outside of the areas described by the four highways. If this trend remains consistent, this Southern Atlantic region where more than 10,000 position reports occur daily would have a mean of fewer than ten that were off highway. This ability to highlight ships that are acting abnormally will better enable maritime analysts to focus their efforts.

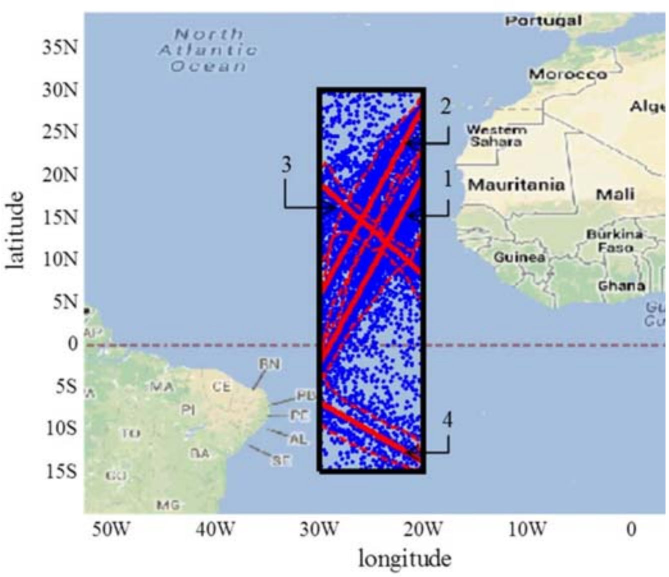

Figure 5. The four most prominent highways in the southern Atlantic Ocean during 2012 are detected by the Hough transformation.

This case study also serves to demonstrate the importance of finding a method to discern traffic patterns that may not be readily apparent in the presence of the most dominant trends. The fourth trend has a ship density of only 0.0953 position reports per month per nm^2, which is lower than the ship density for the region as a whole, which is 0.191 position reports per month per nm^2. This trend might easily be ignored by a method that does not remove the more dominant trends after discovery.

Table 1. Specifications for the four most prominent highways in the southern Atlantic Ocean during 2012 demonstrate how highway ship densities compare.

Rank	% of All Position Reports for Area	Ship Density on Highway (reports per month per nm^2)
1	48.81%	0.5269
2	28.52%	0.3174
3	17.86%	0.3416
4	4.71%	0.0953
Total	99.99%	---

B. *Highway Variability*

The mapping of ocean highways requires periodic update for a variety of reasons. Some are universal to the maritime environment, but others are unique to the type of traffic of interest. Examples include extreme weather, seasonal variations, fluctuation in international markets and trade, and changes in law enforcement monitoring of an area. Seasonal and weather-related fluctuation are universal in impact and offer ground-truth case study opportunities for exploring how the algorithm developed in this research might be used to identify and understand variation in maritime traffic patterns. Conclusions drawn from weather might be adjusted for extension to provide insight into how inside information of market or law enforcement will impact traffic patterns, but these cases are not specifically developed as part of this research.

This traffic analysis method provides a means for assessing the impact of extreme weather patterns on maritime traffic, which can assist in predicting the impact future weather events will have on traffic patterns in the area. Hurricane Ernesto moved on a generally westerly track through the Caribbean in early August of 2012 [6].

The impact of the hurricane can be observed in the atlas generated daily over a four-day period, as displayed in Figure 6. The hurricane track for the day is overlaid in black on top of the highway (in red) detected by the Hough transformation. In this case, a particular highway of interest was selected from all highways found in the area. The highway is mapped using a normalized traffic density of interest of 0.80 and a traffic density tolerance of 0.50 for width determination. Essentially, the southern portion of the highway dissolves as the hurricane crosses its path and then reconstitutes within the next 24 hours as the hurricane moves further west.

Although only a single case study, the persistence of this traffic pattern is evidence in support of the pursuit of an atlas-based anomaly detection method. The ocean may not have lane markings controlling ships' travel, but ships do remain in generally predictable and orderly routes. As soon as the storm had passed, travel along the highway resumed.

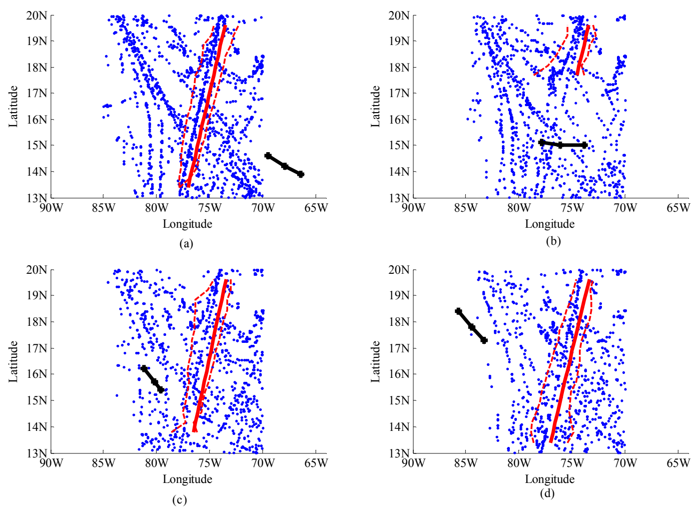

Figure 6. Hurricane Ernesto, plotted in black, caused a highway, plotted in red, to dissolve and then reform as Ernesto passed through the Caribbean over the days of (a) 04-05 August, (b) 05-06 August, (c) 06-07 August, and (d) 07-08 August 2012.

The atlas method developed in this research provides an avenue for gaining insight into normal ship behavior, even in abnormal weather situations. One such behavior pattern is seen in the spike (day 4) in traffic density just before the storm that can be observed along the highway. A plot of traffic density along the highway is presented in Figure 7.

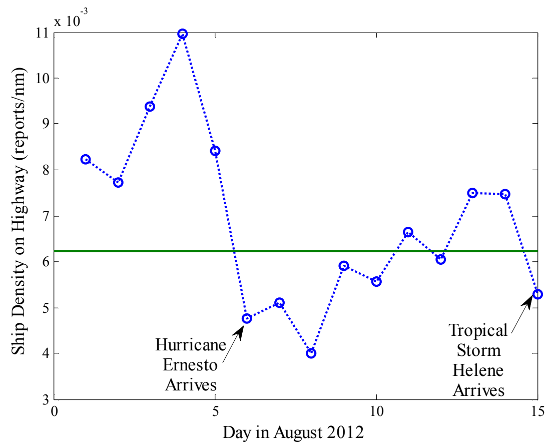

Figure 7. Hurricane Ernesto caused a drop in ship density on an ocean highway during August 2012.

IV. Conclusions

A method for improving maritime domain awareness by developing an atlas of expected ocean traffic patterns was outlined and that atlas was used as the definition of normalcy within an anomaly detection scheme in this paper. A significant contribution of this paper is the application of Hough transformation, an image processing technique to the problem of maritime domain traffic pattern analysis. An atlas generation method was developed beginning with a technique that preprocesses position reports into a traffic density grid. From the traffic density grid, a modified version of the Hough transformation was used to extract highways, which can be compiled and mapped into an atlas of expected ocean traffic patterns. An iterative method was developed to detect less prominent highways. Point-in-polygon problem solving was used to enable geographical region based anomaly detection of ship tracks as compared to the generated atlas. A variety of specific case studies was developed in which the atlas generation and anomaly detection methods demonstrated usefulness to maritime domain awareness. As just one of several examples, the method enabled observations of how a hurricane transformed expected ocean traffic patterns as it traversed the Caribbean. The anomaly detection methods enabled analysis of transit versus normal traffic in a maritime domain scenario.

References

[1] U.S. Department of Transportation Maritime Administration (2011, Feb.). U.S. water transportation statistical snapshot. [Online]. Available: http://www.marad.dot.gov/documents/US_Water_Transportation_Statistical_snapshot.pdf

[2] J. Russ, "Feature Specific Measurements" in *The Image Processing Handbook* (4th ed.). NewYork: CRC Press, ch. 9, pp. 491–496, 2002.

[3] R. Duda and P. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Commun. of the Assoc. for Computing Machinery, Inc.*, vol. 15, no. 1, pp. 11–15, Jan. 1972.

[4] U.S. Coast Guard Navigation Center. (n.d.) *Automatic Identification System overview* [Online]. Available: http://navcen.uscg.gov/?pageName=AISmain

[5] A. McAbee, "Traffic pattern detection using the Hough transformation for anomaly detection to improve maritime domain awareness" M.S.E.E. Thesis, Department of Electrical and Computer Engineering, Naval Postgraduate School, Monterey, CA, Dec. 2013.

[6] D. Brown. (2013, Feb.). Tropical Cyclone Report Hurricane Ernesto (AL052012). [Online]. Available: http://www.nhc.noaa.gov/data/tcr/AL052012_Ernesto.pdf

[7] U.S. Department of Homeland Security (2005, Oct.). National plan to achieve maritime domain awareness. [Online]. Available: https://www.dhs.gov/national-plan-achieve-maritime-domain-awareness